

# Test and Evaluation of Highly Complex Systems

James J. Streilein, Ph.D.

U.S. Army Test and Evaluation Command, Alexandria, Virginia

I have been working in Army test and evaluation (T&E) since 1974 and have seen enormous increases in the complexity of systems since I first started in the business. Systems in 1974 were largely stand-alone, analog, and mechanically controlled. Now to conduct and win the next major conflict with a conventional enemy or against violent extremist movements, previously unimagined systems are being developed and fielded to our warfighters. The complexity of these new systems is a result of addressing today's and tomorrow's threats with more accurate, lethal, reliable, survivable, interoperable, and maintainable systems. Most of today's systems are very software intensive and network enabled, and have on-board, complex subsystems. The complexities that often arise are a result of the interactions among the systems and subsystems, and as a result, they cannot be tested and evaluated in isolation. These systems are vital enablers that assist the warfighters in accomplishing their missions. These new systems are often a system of systems (SoS) on a single platform such as the mine-resistant ambush-protected (MRAP) vehicles or a family of systems such as the Future Combat Systems (FCS), the Stryker family of vehicles, or the Ballistic Missile Defense System.

Testing and evaluating a single service or a joint SoS or a Family of Systems (FoS) requires combat and materiel developers, testers, and evaluators to form larger and more diverse Integrated Product Teams (IPT) and Test and Evaluation Working Integrated Product Teams. These teams must establish and refine the system's requirements, and properly establish and scope the resources and events required to determine the capabilities and limitations of the SoS and/or the FoS under test. In most cases, the SoS or FoS are expensive to produce, train, maintain, and sustain. The cost to properly test and evaluate these systems can run into many millions of dollars. Because of the Service's desire to fill a capability gap and the system's development and production expense involved, many new systems have

increased visibility from the requesting Service and are normally designated by the Department of Defense as a program requiring the oversight of the Director, Operational Test and Evaluation and Developmental Test and Evaluation. This DoD oversight expands T&E complexity, IPT membership, coordination, documentation, planning, resources required, and the program's schedule.

The Army Test and Evaluation Command (ATEC) uses a number of processes and best practices to T&E SoS and FoS per the policy and guidelines directed in the June 2007 Office of the Secretary of Defense Section 231 Report. Some of the challenges for the command in implementing the guidance found in the Section 231 Report are to maximize the use of mission-based test and evaluation, Modeling and Simulation (M&S), joint and distributed testing, reliability growth testing, and determining system interoperability (subsystem, system, and with multiple systems). Let's discuss a few of these challenges.

What is mission-based T&E? Mission-based T&E (MBT&E) is an emerging process to focus T&E of a SoS or a FoS on a system's mission contribution as intended by a combatant commander in accomplishing their assigned mission. This requires the evaluation team's assessment to not only address whether a given system's functionality was sufficiently demonstrated per the critical operational issues and criteria, but to also ascertain for the users and combat developers the likelihood that the SoS or FoS will improve the unit's ability to successfully accomplish their mission. The T&E strategy must do more than check a system's capabilities against the standard type of requirements; now the mission capabilities must also be outlined and a crosswalk developed to ensure that the test events and data will address both system and mission capabilities. However, mission success is often determined by qualitative assessments (military judgment) versus SoS and FoS performance specifications, which are often determined by quantitative data. The determination of whether a system can successfully maneuver



Dr. James J. Streilein

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>MAR 2009</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2009 to 00-00-2009</b>	
4. TITLE AND SUBTITLE <b>Test and Evaluation of Highly Complex Systems</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>U.S. Army Test and Evaluation Command, 4501 Ford Ave, Alexandria, VA, 22302-1458</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>4</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

over a specified type of terrain or operate in a cold or hot environment is easily tested and quantitatively verifiable, but determining whether the system capabilities assist commanders in accomplishing their mission—given the many mission types and threats that exist—is more of a challenge for the T&E community. It is not feasible to test to each mission scenario. Moreover, it will not be practical to test enough replications of the missions and threats to the sample size needed to determine the system's performance with statistical significance. We are led to employ M&S to address mission capabilities, but M&S brings its own level of complexity.

When using M&S to determine system capabilities, its selection as an element of the T&E strategy must also take into consideration that verification, validation, and accreditation must be obtained before M&S is utilized to support an acquisition milestone decision. When M&S is used early in the system's development, it can assist the program manager and contractors to enhance the design. When testing the MRAP, M&S helped to characterize the vulnerability and the survivability of the system. This reduced the time required to develop and test the system.

M&S is a key enabler for effectively focusing and executing T&E. It provides a practical means to support system development, combat development, and T&E throughout complex system program development. M&S helps to prioritize live testing, characterize system attributes, provide information about system performance under conditions that cannot be practically measured with live testing, and reduce overall program risk. Validated M&S expands the test envelope beyond traditional methods required to test today's complex systems. M&S can be used to predict system performance, identify technology and performance risk areas, and support the evaluation of the system's effectiveness, suitability, and survivability.

As we all know, live testing of today's complex systems is becoming increasingly unaffordable. There are many impacts and interactions of adding new equipment to the current force that must be considered and evaluated. In addition, there is limited availability of complementary and adversary systems and forces. M&S surrogates for these systems and forces can effectively flesh out the battle space for live tests. Test range limitations, such as size, availability, cost, security, safety, and environmental concerns must also be taken into consideration. Testing will never be completely physical because of size, complexity, and interoperability requirements that will demand a synthetic environment to be wrapped around the test unit. M&S provides controllable, repeatable testing of components, software, and hardware throughout the

acquisition cycle. M&S can provide a defensible, analytical underpinning for decisions.

The model-test-model approach is often used by ATEC throughout the acquisition life cycle to effectively focus T&E resources on critical test issues. M&S is used to provide early predictions of system performance. Based on those predictions, tests are designed to provide actual data to confirm system performance and validate or accredit M&S. Early in the acquisition and before final configuration hardware is available for testing, M&S can be used to support engineering-level trade studies of technologies and systems and provide data to both the system development and evaluation. Testing M&S can range from computer-based simulations to virtual, wrap-around simulations to hardware-in-the-loop physical testing of components, subsystems, and systems. As hardware matures and becomes available, the evaluation will begin to focus on empirical test data, rather than on the M&S representations. M&S and T&E do not replace each other, they complement each other. The iterative and integrated use of M&S with T&E is of greater value than M&S and T&E conducted in isolation.

Determining system interoperability is also a challenge. Two types of system interoperability must be proven out before fielding: (1) system interoperability where the system's software and hardware can exchange information effectively and (2) the SoS or the FoS exchanging information with existing Army, Joint, and multinational systems and units on the battlefield. The cost of testing interoperability is high. Assembling the architecture needed to test the system's interoperability requires a great deal of hardware and personnel for an extended period. Often testers piggyback on Joint and Service exercises to defray the cost of testing. Using exercises provides test articles and personnel, but testers can lose control of the test event and place their data needs at risk. The primary focus of exercises is not to determine the system capabilities under test or mission success, but to train staff and forces, as well as evaluate plans and strategic operations. Therefore, the scope to determine interoperability must be designed and robust enough so that the service operational test agencies and the Joint Interoperability Test Command can evaluate and certify that the system can generate, deliver, use, and consume data between platforms or systems.

Lastly, the June 2007 Section 231 Report stressed that integrated developmental and operational testing should be used whenever possible to maximize use of all data. This is all the more difficult for complex systems. However, to conduct a single event to collect both developmental and operational data, the event must not conflict with Title 10 independence of

operational testing for materiel developers. If the design and execution of a single test event can support a milestone decision, the test must be engineered so that program managers may still receive an assessment that the system is mature enough to successfully execute missions against realistic threats in operational conditions. A separate event may eliminate the ability to conduct test-fix-test events because the warfighters may not be kept for testing until a fix is developed and applied to the system. Some separate developmental testing (DT) is needed to address the entrance criteria for operational testing (OT) and provide a safety release. Although every effort should be made to conduct DT with operational realism, the IPT may find it difficult to execute only one integrated DT/OT event to evaluate complex systems.

Budget constraints dictate that we make maximum use of T&E resources by combining OT with DT whenever practicable. A single test event for OT and DT has the potential to answer both DT and OT questions efficiently in terms of the time and resources normally required, but it is also the most difficult to execute because it requires maximum coordination and cooperation among members of the test community. When collection of DT and OT measures are integrated, there must be cooperation where all parties stand to benefit. However, more complex systems often have more measures to test and evaluate. The developmental test team, operational test team, and evaluation team must develop a test management structure to share control of the event. When different test agencies participate in the same test event and exchange data, a two-level common language is often required. This language includes terms used to talk about T&E, such as, "issue," "mission," and "measure," as well as language used for evaluating a specific system, such as, "detection," "slant range," and "slant angle." A common language requires standard data definitions and formats while enforcing the specific definition of variables and conditions, and interpretation of results. If the goal of testing is to predict how a system will perform under different conditions, an experimental design must be used that accommodates the needs of both the developmental and operational testers.

For complex systems, the increased number of factors and conditions that are represented across multiple DT and OT data collection phases increase the breadth of the evaluation and the number of questions the evaluator can answer. Metrics that were collected in different event phases (e.g., through both DT and OT) and are complementary to each other might be analyzed together, increasing sample sizes and the confidence of the test results. OT experimental

designs might be designed such that they return DT relevant information and provide useful feedback to the developers. The results of free-play testing should be carefully documented and then analyzed to extract metrics that can be analyzed in concert with, or in addition to, DT metrics. Structuring the factors and conditions such that DT and OT issues are addressed is paramount for a successful test.

We have addressed a few processes and best practices that should be considered when testing a SoS or a FoS system. There are others, but the effectiveness and efficiencies or processes or practices must be explored by the IPT and working integrated product teams.

There are many challenges for the SoS and FoS test and evaluation teams to ensure that they properly determine the system's performance while implementing the principles cited in the DoD Section 231 Report. The teams must remain innovative and use techniques such as MBT&E, cost effective instrumentation, accredited M&S, and optimize resources such as test participants, ranges, and test events. Success in fielding equipment to our warfighters will continue to require total commitment, coordination, and cooperation of all members of the acquisition communities. I have seen the T&E community continually improve over the years since 1974, and I look forward to our efforts and innovations to handle the increases in complexity of systems to be tested and evaluated in the future. □

*DR. JAMES J. STREILEIN entered his current position in April 2007. As the executive technical director/deputy commander, U.S. Army Test and Evaluation Command, he provides oversight and technical direction to all command efforts to include testing, evaluation, modeling and simulation, and instrumentation. He serves as commander in the absence of the commanding general and as the Army member of Technical Advisory Board for Joint Test and Evaluation. He is directly responsible for the Army Quick Reaction Team's work. ATEC is a multibillion dollar command with over one third of all Army land and is currently conducting developmental tests, operational tests, live fire tests, and field data collection on over 400 systems in normal acquisition and over 200 systems in rapid acquisition. He represents ATEC in dealings with program managers, program executive officers, Department of the Army, other services, Joint Improvised Explosive Device Defeat Organization, Office of the Secretary of Defense, Defense Science Board, Army Science Board, National Defense Industrial Association Committee on Operational Test and Evaluation, etc.*

*In September 1999, the Army reorganized test and*



evaluation, and Dr. Streilein was selected as the first director of the newly formed Army Evaluation Center of the Army Test and Evaluation Command. The Army Evaluation Center is the Army's lead for its technical and operational evaluation mission. In the 1996 reorganization of Army test and evaluation, Dr. Streilein was selected as the first director of the Evaluation Analysis Center of the Operational Test and Evaluation Command. Dr. Streilein became a member of the Senior Executive Service in August 1991 upon selection as the

chief, Reliability, Availability, and Maintainability Division of the U.S. Army Materiel Systems Analysis Activity, where he began working in 1974. He received the Presidential Rank Award—Meritorious Executive, 2005; Decoration for Meritorious Civilian Service 1987 and 2007; and Army Superior Unit Awards, 2000 and 2004. He received a bachelor of science degree in mathematics from Carnegie Mellon University and a doctorate in mathematics from Pennsylvania State University.

The European Chapter of ITEA cordially invites you to participate in the



# European Test and Telemetry Conference Conference Européenne des Essais et Telemesure

ETTC'09 theme - Test & telemetry for space and ground transportation

## Toulouse, France 24 - 26 June 2009

Organised by the SEE and 3AF in association with Arbeitskreis Telemetrie EV





For details see: [www.ettc2009.org](http://www.ettc2009.org)

